SENSORY®

VOICEDIRECT ™
Speech Recognition Kit

Now you can add voice recognition to consumer electronics, toys, motors, appliances, lights, games or any electronic product, using award-winning speech technology by Sensory® .... in about an hour!

Kit Contents

- *VOICE DIRECT* ™ Module
- Omnidirectional Electret Microphone Element
- Speaker
- (3) Micro-switches
- (2) 100k resistors
- This Quick Setup Guide

Required Tools and Components

- *Tools:* External 5V(+/-10%) / 50mA power source (equivalent to 3 x 1.5 alkaline batteries), soldering iron, multimeter
- *Helpful components:* 0.1" male headers, 3"x3" proto or perf board, hookup wire

Overview

Sensory's VOICE DIRECT ™ Speech Recognition Kit represents a quantum leap forward in embedded systems speech recognition technology. This kit is an easy-to-use tool for adding voice recognition to a range of applications limited only by your imagination.

VOICE DIRECT ™ is a speaker-dependent speech recognition module, allowing training of up to 15 words. Using sophisticated speech recognition technology, VOICE DIRECT ™ maps spoken commands to system control functions. Each time one of the words is recognized, a corresponding output pin on the module is toggled high for 1 second. With proper use, it can achieve greater than 99% recognition accuracy.

VOICE DIRECT™ Quick Set-up Guide

Contents of this guide:

1. *Introduction*
2. *Features*
3. *Building a Basic Application*
4. *Module Pinout and Characteristics*
5. *Training VOICE DIRECT™*
6. *Module Schematic*
7. *Recognizing Words*
8. *Erasing Templates*
9. *Tips for Optimal Performance*
10. *About Speech Recognition*
11. *Bill of Materials*
12. *How to learn more about VOICE DIRECT™*
13. *Other Sensory® Products*
14. *Warranty*

1. Introduction

This guide will walk you through the steps of bringing speech recognition to your product or application. First, we'll focus on getting you up and running quickly. Later on, we'll discuss tips, tricks, and other information about Sensory® speech recognition technology. For those interested in detailed technical information about VOICE DIRECT™ module and Sensory's award winning speech recognition technology, please visit our web site at www.SensoryInc.com.

2. Features

- Speaker-dependent speech recognition
- Minimal external components
- Recognizes up to 15 words or phrases
- Over 99% recognition accuracy
- Phrase recognition up to 3.2 seconds
- User-friendly speech prompting
- Quick setup time

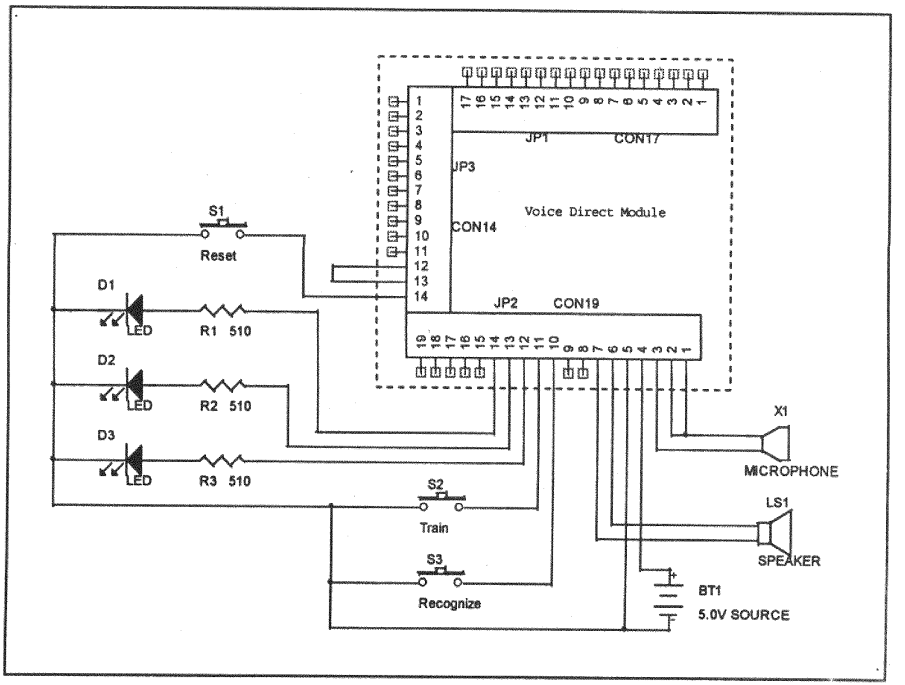
3. Building a Basic Application

Figure 1 shows a schematic for a simple application. There are a few items that should be considered before starting construction:

1) Consider how you are going to fixture the module. Whether or not you mount the module in a box, for example, depends on how much additional circuitry you'd like to add to your application, and how transportable it needs to be.
   a) One possible method of construction is to use a small project box, mounting the speaker, microphone, switches, and module (on standoffs) in the box. For connections, 30 AWG wire-wrap wire can be soldered directly from the top (non-connector side) of the module to the other components.
   b) An alternate method of construction is to use a piece of perf board (approx. 2"x5"), mounting male header strips (0.1" centers) to the perf board to accommodate the module. The perf board can also be used to mount the switches, as well as any other components required by your application. 1/8" mono phone jacks and plugs are useful for quickly connecting and disconnecting the speaker and microphone from the rest of the circuitry.
2) Three 1.5V batteries in a plastic holder can be used as a power source.
3) Consider what types of loads you are going to be driving with the outputs. Most types of loads (LEDs, motors, relays, etc) will require some type of output driver circuitry. A discussion of these circuits is beyond the scope of this guide, but a great place to start looking is in the databooks and on the web pages of semiconductor manufacturers such as National, TI, and Motorola.
4) Remember that the outputs only toggle high for 1 second. If you want them to stay high, you can add a latching output driver.
5) Because VOICE DIRECT ™ can recognize 15 words, but only has 8 outputs, some simple decoding circuitry is required if the application requires more than 8 words to be recognized. The format of the outputted words is discussed under "Recognizing Words" on page 10.
6) Double-check that the microphone element has been connected with the proper polarity. Note that the GND lead of the microphone is connected to its case. The GND lead should be connected to JP2-3 of the module.
7) Use of the 100K resistors included with the kit is optional. See page 5 for a description of using them to set training and recognition selectivity levels.

This sample application will allow the user to train three words and recognize them. When each word is recognized, VOICE DIRECT ™ will turn on the appropriate LED for 1 second. Pressing the Train and Recognize buttons at the same time erases the trained words. Three new words may then be trained.

Figure 1 - ***Sample Application Schematic for Voice Direct ™ Three-LEDs***



4. Module Pinout

Name	Module Pin	Description	Connection	I/O
	JP1 - 1-17 JP2 – 9 JP3 – 0-9	Unused	NC	-
PREAMP IN	JP2 - 1	Microphone Input Connection	Connect to microphone, other microphone connection to GND	I
MIC BIAS	JP2 - 2	Mic Bias (Electret microphone)	If powered mic. being used, NC, otherwise connect to JP2-1	I
AGND	JP2 - 3,5	Analog Ground. For noise reasons, analog and digital grounds should connect together only at VOICE DIRECT ™ (VDI)	GND	-
+5V	JP2 - 4	5 Volt (+) Power Supply Connection	VCC	-
PWM1	JP2 - 6	Pulse Width Modulator Output 1 (multiplexed)	Connect to 8 to 32-Ohm speaker. Provides approximately 0.15 Watts of audio power in to 32-Ohms	O
PWM0	JP2 - 7	Pulse Width Modulator Output 0	Connect to 8 to 32-Ohm speaker. Provides approximately 0.15 Watts of audio power in to 32-Ohms	O
DACOUT	JP2 - 8	Analog Output (unbuffered).	High-impedance (22kOhm) analog audio output. Must be power amplified to drive a speaker, and should be low-pass filtered with a corner frequency around 20 KHz. Better speech quality than the PWM output. Recommended for applications requiring either louder volume or better speech quality.	O
- RECOG	JP2 – 10	Recognition sensitivity selection and activate recognition	*See below*	I
- TRAIN	JP2 – 11	Training sensitivity selection and activate training	*See below*	I
OUT1 – OUT7	JP2 – 12-18	Stand Alone mode output ports 1 -7	Connect to user application's control lines.	O
HIGH/ OUT8	JP2 – 19	Stand Alone mode output high or output port 8	Connect to user application's control lines.	O
ERROR	JP3 – 10	Stand Alone mode error signal	Connect to user application's error indication lines.	O
GND	JP3 - 11,12	Digital Ground, CPU core (pins 1 and 33) and I/O (pins 18 and 52)	GND	-
MODE	JP3 – 13	Stand Alone or Slave mode selection.	100K Resistor to GND	I
-RESET	JP3 - 14	Reset	VCC	I

The -TRAIN and –RECOG pins control the selectivity and activation of VOICE DIRECT ™. The training and recognition selectivity levels are set when VOICE DIRECT ™ is first powered up (or reset) according to the following table:

Pin Configuration	Selectivity
-TRAIN open circuit	Relaxed Training – Easier to train, accepts more similar sounding words (fewer rejections).
-TRAIN pulled to GND with a 100K Resistor	Strict Training – Harder to train, rejects more similar sounding words – may result in better recognition accuracy (fewer substitutions).
-RECOG open circuit	Relaxed Recognition – Recognizes more words, may substitute the wrong words (fewer rejections).
-RECOG pulled to GND with a 100K resistor	Strict Recognition – Recognizes fewer words, may reject trained words (fewer substitutions).

To start training, pull the –TRAIN line to GND for at least 100mS. To start recognition, pull the –RECOG line to GND for at least 100mS. To erase all recorded words, pull both the –TRAIN and –RECOG pins to GND for at least 1S.

Module Pin Characteristics

Absolute Maximum Ratings

Minimum voltage on any pin	V_{ss}-0.6V
Maximum voltage on any pin	V_{dd}+0.6V
Operating temperature (T_O)	-20°C to +70°C
Soldering temperature	260°C for 10 sec
Maximum voltage	7.5V
Power dissipation	1W
Storage Temperature	-65°C to +150°C
Minimum Operating Voltage	3.5V
Maximum Operating Voltage	5.0V

WARNING: Stressing VOICE DIRECT ™ beyond the "Absolute Maximum Ratings" may cause permanent damage. These are stress ratings only. Operation beyond the "Operating Conditions" is not recommended and extended exposure beyond the "Operating Conditions" may affect device reliability

DC Characteristics

(T_O = -20°C to +70°C, Vdd = 5V)

SYMB	PARAMETER	MIN	TYP	MAX	UNITS	TEST CONDITIONS
V_{IL}	Input Low Voltage MODE, I/O pins RESET	 -0.1 -0.1		 0.75 0.60	 V V	
V_{IH}	Input High Voltage MODE, I/O pins RESET	 2.5 3.0		 Vdd+ 0.3 Vdd+ 0.3	 V V	
V_{OL}	Output Low Voltage I/O pins			0.5	V	I_{OL}= 2.0 mA
V_{OH}	Output High Voltage I/O pins	4.0			V	I_{OH}= -2.0 mA
I_{IL}	Input Leakage Current MODE, I/O pins RESET		 <1 <1	 5 5	 uA uA	 $V_{ss}<V_{pin}<V_{dd}$ $V_{ss}<V_{pin}<V_{dd}$
I_{cc1}	Supply Current, Operating		30	100	mA	
I_{cc2}	Supply Current, Quiescent		20	30	mA	
R_{PU}	Internal Pull-up Resistance, I/O pins		4.5, 200, Hi-Z		k Ohms	Software selected

Analog Characteristics

SYMB	PARAMETER	MIN	TYP	MAX	UNITS	TEST CONDITIONS
V_{DO}	DACOUT Voltage Swing	Agnd		Avdd	V	

5. Training VoiceDirect™

When the –TRAIN pin is pulled to GND for at least 100mS (such as by pushing a momentary switch), training will begin. VOICE DIRECT ™ will prompt "Say word *x*" (where *x* corresponds to the word to be trained).

A word or phrase must be shorter than 3.2 seconds and may not contain silences longer than 0.5 seconds. For example, the name "John Smith" would be an acceptable phrase as long as the two words are not separated by a large pause. Training terminates when no word is spoken in response to a prompt, when either switch is pressed a second time during training, when three errors have occurred during training or after all words have been trained.

Pressing the TRAIN switch at a later time resumes training. New words are added to the end of the set already recorded. New words may be added to the set at any time, up to a maximum of 15 words. Individual words from the set may not be deleted or overwritten, but the entire set can be erased (see "Erasing Templates" Section 8). The user says the first word to be trained, then VOICE DIRECT ™ prompts again with "Repeat". The user repeats the word, and VOICE DIRECT ™ will return "Accepted" if the word has been successfully trained, otherwise it will say the cause of the training error. If an error occurs during training, then the error will be spoken "Spoke too soon", "Please talk louder", etc., and the ERROR line will pulse high for 1 second. The user will get three attempts to train each word before VOICE DIRECT ™ exits training mode, says "Training complete", and pulses the ERROR line high for 1S.

The user can exit training at any time by pressing the –TRAIN or the –RECOG button, by not responding to a "Say word *x*" or "Repeat" prompt, or when all 15 words have been trained.

6. VoiceDirect™ Schematic

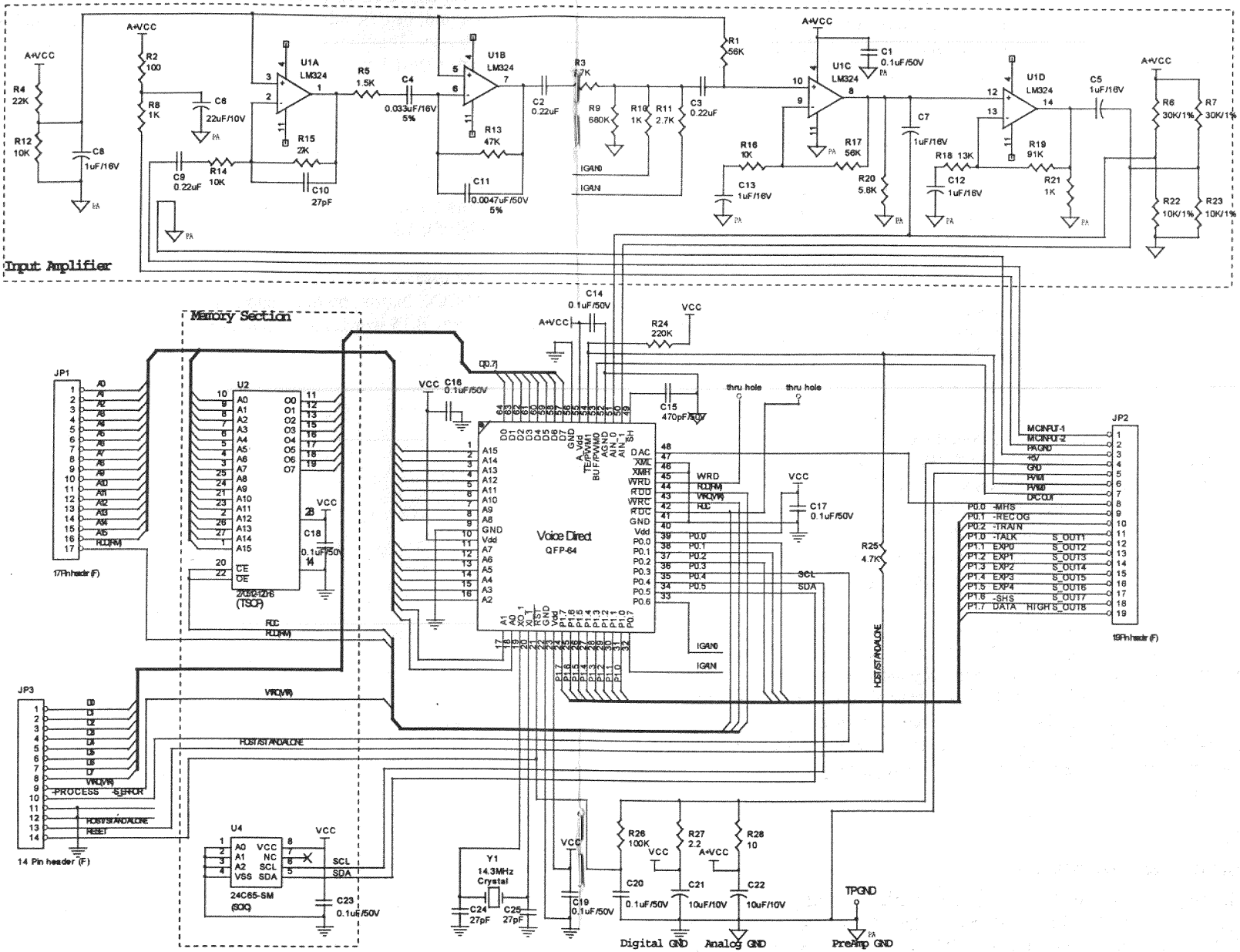


Figure 2

7. Recognizing Words

When the –RECOG pin is pulled to GND for at least 100mS, recognition will begin. VOICE DIRECT ™ will prompt "Say a word". If the response is not recognized, VOICE DIRECT ™ will say "Word not recognized" and exit recognition mode. When a trained word is successfully recognized, the appropriate Output Pin(s) will pulse high for 1 second:

Word 1	Output 1
Word 2	Output 2
Word 3	Output 3
Word 4	Output 4
Word 5	Output 5
Word 6	Output 6
Word 7	Output 7
Word 8	Output 8
Word 9	Output 8 and Output 1
Word 10	Output 8 and Output 2
Word 11	Output 8 and Output 3
Word 12	Output 8 and Output 4
Word 13	Output 8 and Output 5
Word 14	Output 8 and Output 6
Word 15	Output 8 and Output 7

Special Note: Because the module can recognize 15 words, but only has 8 output pins, words 9 through 15 are represented in binary coded form, as show in the table to the left. If an application requires that a unique pin pulse high for each of the 15 words, a decoding circuit should be added to the output of these

If an error occurs during recognition (except for "Word not recognized"), then the error will be spoken "Spoke too soon", "Please talk louder", etc., and the ERROR line will pulse high for 1 second. If the spoken word is not recognized for any reason, none of the pins is activated and an appropriate voice message is synthesized. Since this can occur if VOICE DIRECT ™ has not been trained on the word, the "word not recognized" output is not considered an error condition, and the ERROR bit remains inactive.

VOICE DIRECT ™ monitors the background noise level and gives a voice warning if the noise is too high for recognition. VOICE DIRECT ™ works well with high (above 80 dB) levels of *steady* background noise (such as a fan), but it may make errors at lower levels if the background noise is not steady (for example, a TV set). Best results will be obtained in a relatively quiet location. Warnings may also appear if the word is spoken too softly, too loudly, or too quickly after the prompt. The ERROR signal is toggled high when these conditions occur.

8. Erasing the Templates

When both the –TRAIN and –RECOG pin are pulled to GND for at least 100mS, VOICE DIRECT ™ will erase all trained templates. VOICE DIRECT ™ will say "Memory erased", pulse the ERROR line high for 1S, and then exit the erase mode.

9. Tips for Optimal Performance

Successful recognition begins with the careful selection of words for each recognition set. Several factors contribute to selecting an optimal recognition set. Problematic recognition sets can often be corrected by replacing one or more words with a synonym, or approximate synonym (see examples, below), without requiring any other changes. The smaller the set, the higher the recognition rate.

The optimal set consists of:	Avoid sets like:	Aim for sets like:
Dissimilar sounding words	hat/cat/rat home phone/office phone	hat/kitten/mouse home/office
Varying numbers of syllables	orange/apple/cherry	orange/watermelon/ grape

Key considerations for successful voice recording

- The equipment used to train the voice recordings should match the equipment used during recognition. Differences in microphone, microphone housing, etc. will adversely affect recognition.
- The conditions and environment in which the voice recordings are made should reflect the conditions and environment in which the end product will be used.

a) **Distance.** The distance of the microphone from the speaker's mouth must be the same during recording and during end-product use. For example, a doll is typically held within arm's length, and so the voice recordings microphone should be held accordingly.

b) **Natural Voices.** Subjects should speak in their normal voice and should be discouraged from sounding different by imitating a foreign accent or using any unnatural intonation. They should be prompted by means of some non-verbal source (pictures or flash cards, for example), so as not to unconsciously mimic the voice of the person doing the prompting.

c) **Physical States.** Physical states should be considered. For example, in collecting voice recordings for an exercise machine, it is strongly recommended to record people who are breathing heavily.

d) **Emotional States.** Emotional states should be considered. Will the end users be relatively quiet and calm (e.g., for an office product) or loud and excited (e.g., for a playground toy)?

e) **Environment / Background Noise.** Environmental noise must be considered. Voice recordings should ideally be made in an environment similar to the one in which the end product will be used. For a product intended for use in cars, it is advisable to record in a running, moving car. Nevertheless, the recordings should preferably be made in a reasonably quiet environment. The speech signal must be prominent relative to background noise and there should not be any abrupt, loud noises. Voice recordings should not be made in a soundproof room. These rooms lend an unnatural background silence to the recordings which does not reflect the real-world environment in which the end product will be used.

10. About Speech Recognition

VOICE DIRECT ™ performs speaker-dependent discrete word recognition by comparing a pattern generated in real time with previously trained word templates. The pattern generated by VOICE DIRECT ™ is based on a digital reconstruction of the voice command. Each word to be recognized must first be *trained.* During training, VOICE DIRECT ™ constructs a *template* representing the individual speaker's unique sound pattern for each specific word or phrase to be recognized. Templates are stored in serial EEPROM memory. During recognition a new pattern is produced and compared to the stored templates to determine which word was spoken. VOICE DIRECT ™ features integrated speech-prompting for both training and recognition operations, allowing the development of sophisticated interactive products with minimal programming. VOICE DIRECT ™ performs the following operations when recognizing a word:

1. The audio signal (spoken word) is externally amplified and filtered and then supplied to the analog inputs of VOICE DIRECT ™, which converts the analog waveforms to digital samples.

2. VOICE DIRECT ™ analyzes the speech signal samples and generates a pattern of information representing significant speech elements.

3. VOICE DIRECT ™ increases or decreases the gain of the external amplifier as needed to maintain signal quality.

4. Using a neural network, the pattern is compared with previously stored template patterns; a small number of candidate templates is selected.

5. The candidate templates are further processed to determine the one template that provides the best match to the unknown pattern.

6. If the best match template gives a score above a pre-defined threshold, VOICE DIRECT ™ chooses the word associated with that template. If no template provides a match above threshold, a special "no match" value is chosen.

Steps 1-3 (prior page) are repeated for each word during training. VOICE DIRECT ™ stores the average of two training patterns for each word to improve accuracy. Before storing a new template it is compared with the existing templates in the set. The new candidate word will not be accepted if it is too similar to an existing word (for example, "Bill Smith" and "Jill Smith").

VOICE DIRECT,™ like other speech recognition systems, is necessarily subject to two types of errors: *rejects* (failure to recognize a word in the vocabulary) and *substitutions* (confusion of two vocabulary words, or recognition of a non-vocabulary word). The relative importance of each type of error may depend on the application. VOICE DIRECT ™ provides selectivity levels that allow the user to optimize the tradeoff between these two types of recognition errors. When the recognition selectivity level is set to its highest value, VOICE DIRECT ™ minimizes substitution errors but may produce more rejection errors. When the training selectivity level is set to its highest value, VOICE DIRECT ™ minimizes both substitution and rejection errors by rejecting words that are too similar sounding, potentially increasing recognition accuracy. These settings are best established by experimenting with your own application and environment.

11. Bill of Materials for VOICE DIRECT ™

Qty	Reference	Part
8	C1,C14,C16,C17,C19,C20, C23,C18	0.1uF 50V
3	C2,C3,C9	0.22uF 16v
1	C4	0.033uF 16V
5	C5,C7,C8,C12,C13	1uF/16V
1	C6	22uf/10v
3	C10,C24,C25	27pF 50v
1	C11	0.0047uF 50V
1	C15	470pF 50V
2	C22,C21	10uF/10V
1	JP1	17P F header
1	JP2	19P F header
1	JP3	14P F header
2	R1,R17	56K
1	R2	100
2	R3,R25	4.7K
1	R4	22K
1	R5	1.5K
2	R6,R7	30K/1%
3	R8,R10,R21	1K
1	R9	680K
1	R11	2.7K
3	R12,R14,R16	10K
1	R13	47K
1	R15	27K
1	R18	13K
1	R19	91K
1	R20	5.6K
2	R23,R22	10K/1%
1	R24	220K
1	R26	100K
1	R27	2.2
1	R28	10
1	U1	LM324M
1	U2	27C512
1	U3	RSC-164v2
1	U4	24C65-SM (SOIC)
1	Y1	14.3MHz Crystal

12. How to learn more about VOICE DIRECT™

Voice Direct™ has more advanced capabilities designed for embedded systems developers. For a detailed databook that includes detailed technical specifications and programming information on using Voice Direct™ as a slave-mode voice recognition processor, visit our website at www.SensoryInc.com.

13. Other Sensory® Products

THE INTERACTIVE SPEECH™ PRODUCT LINE

The Interactive Speech line of ICs and software was developed to *"bring life to products"* through advanced speech recognition and audio technology. The Interactive Speech Product Line was designed for consumer telephony products and cost-sensitive consumer electronic applications such as home electronics, personal security, and personal communication. The product line includes award-winning RSC-series general purpose microcontrollers plus a line of easy-to-implement chips which can be pin-configured or controlled by an external host microcontroller. Sensory's software technologies run on a variety of microcontrollers and DSPs.

RSC Microcontrollers

The RSC family of microcontrollers (RSC164, RSC264T, RSC364) are low-cost 8-bit microcontrollers designed for use in consumer electronics. All members of the RSC family are fully integrated and include A/D, D/A, ROM, and RAM circuitry on chip. The RSC264T and RSC364 also include on-chip pre-amplification. The RSC family of microcontrollers can perform a full range of speech/audio functions including speech recognition, speaker verification, speech and music synthesis, and voice record/playback

Voice Direct™ TSSP

Voice Direct™ TSSP provides cost-sensitive products with speaker-dependent speech recognition and speech synthesis. This easy-to-use, pin-configurable chip requires no custom programming and can recognize up to 60 trained words. Voice Direct™ TSSP is most ideal for consumer telephony products which feature voice dialing.

Voice Dialer™ ASSP

The Voice Dialer™ ASSP delivers speech recognition technology that allows users to dial phone numbers by saying the name of the person they wish to call. Voice dialing and phone directory management through speech recognition can be easily integrated into existing products. This IC, which includes DTMF tone generation, is designed for use as a slave chip controlled by an external host processor.

Voice Dialer Software

The Voice Dialer software provides advanced speech technology on a variety of microcontroller and DSP platforms. A complete speech API and flexible design allows manufacturers to easily integrate speech functionality into telephony products.

14. Voice Direct™ Speech Recognition Kit Limited Warranty

The Voice Direct™ Speech Recognition Kit is warranted against defects and workmanship for a period of 90 days from the date of product purchase. Sensory, Inc. will, at its option, either repair or replace a product that proves to be defective either upon receipt or through normal usage. If a Sensory® Speech Recognition Kit product has been obsoleted or is no longer in production and deemed non-repairable, Sensory will, at its option, provide an equivalent product or system for a nominal fee.

Sensory ® warrants this Speech Recognition Kit product, when properly installed and used, will execute its programmed instructions. However, Sensory, Inc. does not warrant that the operation of the Product, its firmware and software will be uninterrupted or totally error free. The Product must be returned to Sensory, Inc. for warranty service within the warranty period to the following address: Sensory, Inc., 521 E. Weddell Drive, Sunnyvale, CA 94089-2164. The Buyer will pay all shipping and other charges or assessments for the return of the Product to Sensory, Inc.

Limitation of Warranty

The foregoing warranty shall not apply to defects resulting from maintenance performed by anyone other than Sensory, Inc., modifications made by Buyer or any third party, Buyer supplied software or interfacing, misuse, abuse, accident, mishandling, operation outside the environmental specifications for the Product, or improper setup or maintenance.

Limitation of Liability

Sensory's liability shall be limited to the repair or replacement of defective products in accordance with the Voice Direct Speech Recognition Kit Limited Warranty. Sensory shall not be liable for any incidental, special or consequential damages for breach of any warranty, expressed or implied, directly or indirectly arising out of Sensory's sale of merchandise, including any failure to deliver any merchandise, or arising out of customer's installation or use, whether proper or improper, of the product, separately or in combination with other equipment, or from any other cause. Products sold by Sensory are not authorized for use as critical components in life support devices or systems.

Exclusive Remedies

The remedies provided herein are Sensory's sole liability and Buyer's sole and exclusive remedies for breach of warranty. Sensory shall not be liable for any special, incidental, consequential, direct or indirect damages, whether based on contract, tort, or any legal theory. The foregoing warranty is in lieu of any and all other warranties, whether express, implied, or statutory, including but not limited to warranties of merchantability and suitability for a particular purpose.

IMPORTANT NOTICES

Sensory reserves the right to make changes to or to discontinue any product or service identified in this publication at any time without notice in order to improve design and supply the best possible product. Sensory does not assume responsibility for use of any circuitry other than circuitry entirely embodied in a Sensory product. Information contained herein is provided gratuitously and without liability to any user. Reasonable efforts have been made to verify the accuracy of this information but no guarantee whatsoever is given as to the accuracy or as to its applicability to particular uses.

Applications described in this data sheet are for illustrative purposes only, and Sensory makes no warranties or representations that the RSC series of products will be suitable for such applications. In every instance, it must be the responsibility of the user to determine the suitability of the products for each application. Sensory products are not authorized for use as critical components in life support devices or systems. Sensory conveys no license or title, either expressed or implied, under any patent, copyright, or mask work right to the RSC series of products, and Sensory makes balance between recognition and synthesis no warranties or representations that the RSC series of products are free from patent, copyright, or mask work right infringement, unless otherwise specified. Nothing contained herein shall be construed as a recommendation to use any product in violation of existing patents or other rights of third parties. The sale of any Sensory product is subject to all Sensory Terms and Conditions of Sales and Sales Policies.

© 1999 SENSORY, INC.
ALL RIGHTS RESERVED
P/N 80-0124-B

Sensory® is registered by the U.S. Patent and Trademark Office. Sensory® is a registered trademark of Sensory, Inc. All other trademarks or registered trademarks are the property of their respective owners.